Data Warehouse and BI to Catalize Information Use in Health Sector for Decision Making: A Case Study

Selemani Slum Ally
Liverpool University, UK
selemani.ally@online.liverpool.ac.uk

Nawaz Khan
Department of Computer Science
Middlesex University, UK
n.x.khan@mdx.ac.uk

***Abstract*— Health sector in most countries including Zanzibar lacks the single repository that integrates health management data produced by multitude of sources. The study proves cost effective but successful implementation of Data Warehouse (DW) plays useful role to influence informed decision making in the health sector. Successful DW needs to include necessary features such as Online Analytic Process (OLAP) which gives decision makers the freedom to analyses information into different format business intelligence (BI) tools, data mining and data presentation tools. In this study, we attempted to identify DW needs to consider culture of organization and what organization actual wants for decision making. The study proposed the cost effective DW and BI tools solution to catalyze informed decision making in health care industry. The study also aims to discuss challenges and sets out recommendations.**

Keywords—DW, BI, Health Sector, Decision Making. Zanzibar

I. INTRODUCTION

The health sector is the sector that collects huge volume of data routine and non-routine. Non routine data includes health researches, studies and surveys. Unfortunately, level of informed decision is considerably low in health sectors [1], still many decision makers are making decisions in the "old-fashioned way" using a lot of "gut instinct" [2]. As part of intervention the public health world is making much efforts to foster informed decision in the health sectors [3]. Decision makers need information that makes sense and is actionable, but most health systems in the world "fails to fully link evidence to decisions and suffer from a reduced ability to respond to priority needs at all levels of the health system"[4]. Edward T. Chen [15] refers to healthcare system as 'data rich' but 'knowledge poor'. This is absolutely correct as the health sector collects lot of data. Lack of proper single repository leads data being stored in informal stand-alone application like EPI-info, access or excel. As a result, the collected data failed to produce knowledge for informed decision making. There are a number of factors contributing to the low level of informed decisions. Main obstacles that hinder the decision making process in health organization, the factors mentioned by both, are [4 and 15]:

1. The lack of data ownership among managers,

2. Limited knowledge about usefulness of data, analysis and use of data on management,

3. Perception and attitude that collected have poor quality,

4. Absence of accessible system that does not make data user-friendly and on demanded aggregation

A Data Warehouse (DW) is indeed the solution proposed to mitigate the problem by ensuring that "appropriate data is available to the appropriate end- user at the appropriate time" [5].Despite the improvement in the level of informed decision achieved by other organizations via DW, and efforts shown by leading international organizations for devising DW in health sectors, success remains few & elusive and challenging to achieve [6] Most of DW projects have inherent risks and are prone to failure due to several challenges [7]. Implementation of a successful DW is expensive and requires considerable resources and much efforts from technical and non-technical personnel [8]. Data integration process, (alias ETL) come from different databases designed on different architectures and data quality [9] and therefore the rate of DW adaptation in health sectors is considerably low.

Several researchers developed DW models. Unlike other researchers, the objective of this study is not "how to" design DW but "what if" you have DW in place as far as informed decision making in health sectors is concerned. For effective use of DW one needs more than DW, which just stores data extracted from different sources. Optimum results of DW need adherence to both discipline and best practice. Efficiency of DW may be achieved if organizations implement the complete sets of DW and Business Intelligence (BI) tools with combination of technology (architecture and tools), capable of integrating different format of data coming from different sources, query to perform data mining, analysis and presentation of information in proper format and demand aggregation [9] and Online Analytic Process (OLAP) functionality.

Previous studies show that scorecard is among the dominating tools in measuring and improving the performance of specific health care areas [10]. DW and BI are both experiencing huge demand in the health sector as the sector finds it valuable [11]. Proper development, successful implementation and efficient uses of DW and BI will lead to sensitizing informed decision making. It is important to note that both tools need to be "more science than art" [9].

This paper is organized as such: section 2 presents the general concepts of DW and associated tools, section 3 presents research methodology; section 4 presents the DW for Zanzibar; section 5 presents the evaluation, challenges and recommendation and finally the conclusion of the paper is presented in section 6.

II. DW, BI AND DATA MINING IN FOSTERING DECISION MAKING

A. Data Warehouse (DW)

A DW is another level of database in management information system which acts as "One stop shopping" [11] focusing on supporting informed decision making. A DW is a centre for knowledge creation, which mitigates challenges of having multiple sources of data , which are not speaking to each other [12]. DW should be maintained separately from operational databases [10]. Unlike operational database, DW collects consolidated and summarized data that are useful in decision making process. The decision makers need right information at right time in right format at the right location [2], so does DW. Successfully implemented, DW can directly provide necessary knowledge for supporting decision makers to take the best decision for improving the health of population [13]. Data warehouse's standard disciplines has four essential characteristics, proposed by two DW icons known as Kimble and Inmon, "Integrated, Subject-oriented should, Time-variant, Non-volatile" collection of data [14].

B. Business Intelligence (BI)

DW and BI always complement each other in that classic notion that says "I scratch your back and you scratch mine". BI in a big picture can be defined as a "framework that allows a business to transform data into information, information into knowledge, and knowledge into wisdom" [14]. BI is all "process of getting the right information to the right decision makers at the right time and in the right format" [16]. Producing meaningful information, which can easily be digested or transformed into sending signs of strength and weakness of health system to decision makers is the main objective of BI. DW and BI tools are the best friends for life and their combinations are excellent[2]. DW without BI remains "underutilized and untapped" [15].

C. BI Tools

BI needs to show clear insight into where the performance was in the past, where it is today and where it is heading in future. Dashboard and Scorecard are two BI tools which are relatively new in health sector though for years these have been utilized in other sectors. Balanced scorecard is a management system that used to monitor due performance of strategic goals of organization using color cord to identify target cut-off points. Organizations use scorecard to align their business activities towards their vision and strategy [17]. Scorecard amazingly gets real interest into health sectors African Leaders Malaria Alliance (ALMA) is good example of organizations that foster and promote the effective usage of scorecards in health sector.

D. Data mining and presentation tools

Data mining can be defined in different ways. However, all definitions agree that it is a process of extracting sensible information out of pool of data. Data mining tools are useful to retrieve relevant knowledge from the large repository of data [18]. Sensible Information is indeed source of all knowledge. Basically, it must answer the common three pillars of the three mandatory requirements -WHAT, WHEN and WHERE. The data mining tools should work with OLAP to enable fetching and manipulating data into disaggregation and aggregation, as demanded by decision makers. The output of information should be presented in the format that gives quick and self- elaborative snapshot to decision makers in the graphic presentation like Pivot tables, Geographic Information System (GIS), Graphs, and Charts, which create conducive environment for interpretation of information for better planning and sound decision. Emergence of GIS in health is an added advantage, especially knowing where things happen. It provides knowledge concerning the catchment population that the sector serve [8].

E. OLAP

DW becomes more effective if it supports OLAP [8]. OLAP speeds up online data process whereby query results come within low latency. Some decision makers may want to do deep diving on data analysis. Therefore, aggregation or disaggregation (drill-up and down) to specific hierarchy of management in order to identify problem may be necessary. The best OLAP should support five types of queries named by: "slice and dice, drill down and roll up, drill across, ranking and rotating" [10].

III. RESEARCH AND METHOD

The research methodology is based on Case study and Action Research which was conducted in Zanzibar health sector. Action research is best defined as an initiative which is not limited to problem solving only but also to lead to recommended guidelines for best work practice in organizations. The Health information system (HIS) survey, conducted in 2011 in Zanzibar by WHO [12] revealed three main challenges in HIS, which as a result hinders informed decision-making. There is lack of integration between the existing Health information systems, lack of integrated Decision Support tools and Insufficient uses of information in planning and decision making (Zanzibar HIS, 2011). Health Key performance indicators (KPIs) determine the sources of information to be integrated.

The DW study based on the World Health Organization (WHO) model proposed the six building blocks of health consolidation and modernization. Each of these blocks has KPIs to monitor, manage and plan for the performance of the health sector. The blocks are:-

i. Service deliveries
ii. Health Workforce (Human resource)
iii. Health information System (HIS)
iv. Access to Essential Medicine
v. Health Financial
vi. Governance and leadership

The objectives of Methodology for the project were divided into main five (5) areas:-

i. Investigation of existing operational databases for Health information system (HIS) on interoperability,

security and transfer of selected information or Extraction, Transformation, and Loading (ETL)

ii. Assessing Data Demand and Use (DDU) to obtain the list of KPIs, milestones, targets of sector.

iii. Implementing an effective, efficient and cost-effective (Low budget) DW prototype with suitable Decision Support tools based on best- practice Health Business Intelligence (BI) tools worldwide

iv. Evaluation of existing historical data over the period of 10 years and imported to the DW.

v. Testing and Evaluation of DW by filling the questionnaires

The DW methodology governs the design of DW; it has four IT artifact stages named as Integration, Implementation, Intelligence and Innovation [20].

Two questionnaires asked range of questions from decision making to technical issues. Both collected quantitative and qualitative data-probing from the standard of data warehouse, customer satisfaction as far as decision making process is concerned, data mining and presentation, BI tools, Data Manipulation, usefulness of OLAP, return on Investment (ROI), flexibility, accessibility, performance, risks and challenges etc. A total of 12 IT experts from different government sectors participated in technical evaluation. In total 38 staff, working in the Zanzibar health system, participated in decision making evaluation process.

IV. Implementation

The DW solution is based on free and Open Source software known as DHIS2 [20], which stands for District Health Information Software developed. The software, developed and supported by University of Oslo stands under GNU license.

A. Data Warehouse Architecture

DW is developed from DHIS2 software model. Therefore it inherited the architecture, Data Models, Dimensional models and attributes hierarchy. DW solution is centralized and uses single data-mart. DHIS2 was developed with inter-operability concept in mind to harmonize data from different sources. The data-mart boosts the performance of OLAP as it aggregates reports at different levels of hierarchy and periods in advance (scheduled as nightly jobs). The DW integrated data from Zanzibar Health Management Information System (HMIS), Human Resource Data (iHRIS), Financial, Logistic Management System (eLMS) and from different Surveys and studies (Fig.1). The prototype DW target strategic and tactical decision makers.

B. Interoperability (ETL)

Integration Process conducted into three steps known as Extract, Transform and Load (ETL) process. DW supports Comma Separated Values (CSV), XML and JSON data format. The projects applied basic (unsophisticated) ETL process using CSV. The excel application acted as Data staging area (DSA), whereby mapping activities of data elements, period, hierarchy are duly conducted. The data extracted from the data sources above, need to be transformed, mapped, cleansed and quality checked before being imported to the DW. Data value model has three characteristics- Data element, Period by which event occurred and Source (Organization unit).

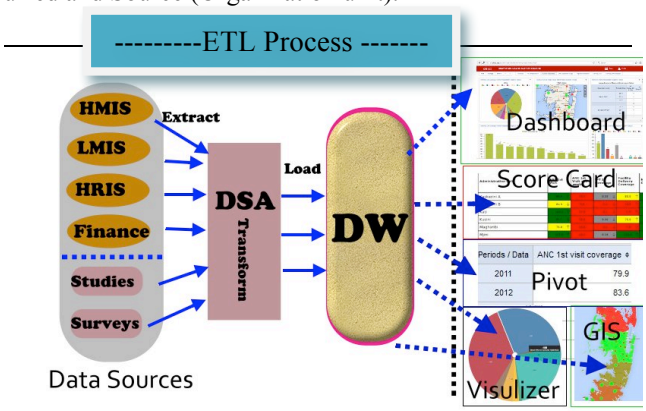

Fig. 1. Data Warehouse Architecture

Any data value (Fig. 2) answering the three categories can be imported into DW. Below is the diagram showing three characteristics that should be possessed by data value, dimensions and facts, star scheme (Fig. 3).

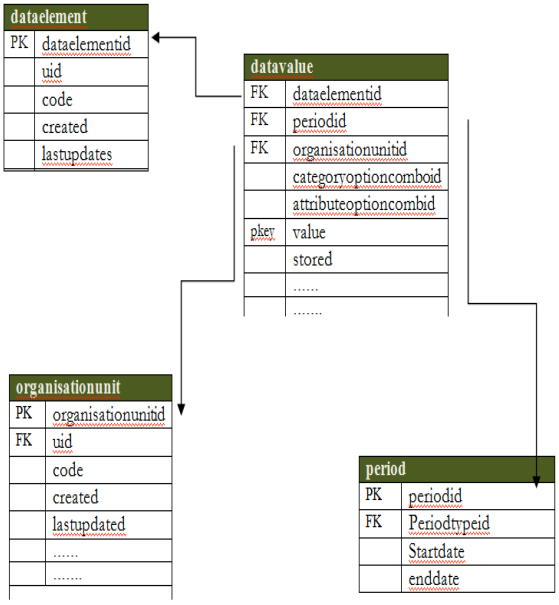

Fig. 2. Data model of data value

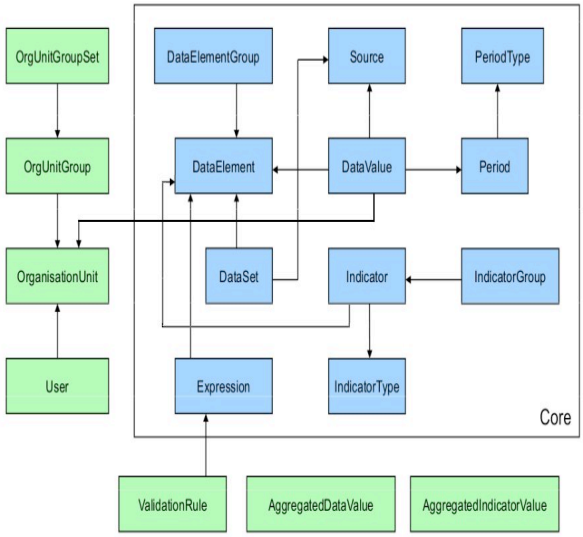

Fig. 3. Simplified Star Schema for Zanzibar HIS

V. Evaluation, Findings and Discussion

This study proves that cost effective but "successfully" implemented DW and BI solution catalyze informed decision-making in health sector. The words successfully implemented cover all process of DW including business requirements, development, DW criteria, data mining functionalities and BI tools. Successfully implemented DW goes beyond the technical evaluation; end-user satisfaction is critical success for DW implementation. [4] cited that: "Measurement of the success of the Data Warehouse is outside the experience of information technology specialists". Other non-technical factors are based on best practice, cultural accessibility, sustainability and data quality. The study revealed that 100% of Decision makers found online DW improve accessibility and availability to eligible users anywhere, anytime.

DW Cost effectiveness may be defined as development of efficient DW and BI tool but within reasonable and affordable amount of cost. The participants voted the solution to have low total cost of ownership (TCO) in term of technology and other resources i.e. money, time, human, etc. This is another essential factor especially for resource- constrained countries like Zanzibar. Furthermore, it can easily be accessible online, manageable and scalable leading to optimum performance, usability and efficiency.

The decision makers supported the idea of having DW and BI focusing unselected lists of key performance indicators (KPIs). Collecting KPIs from multitude of fragmented HIS systems is time consuming and not user-friendly to decision makers. DW simplifies process, reduces the efforts and save times that can effectively be spent for data analysis and interpretation of data, brainstorming on planning decisions. These are strong advantages in catalyzing informed decision and lead to interventions based on knowledge acquired. Otherwise the absence of DW may results into low -data ownership in organizations. The decision makers could be unaware of existing relevant data for decision -making [18].

Health workers travel a lot from one geographical area to the other for meetings, seminars etc. The world is experiencing very rapid changes in handheld technology' it is an advantage that a big percentages of employees own and spend most of their time with smart phones, tablets or other hand-held items. The research shows that 81.6% of decision makers are eager to access DW's dashboard, using their handheld devices to self-service data whenever they need. Therefore, the DW needs to be responsive or adaptive to hand-held devices. The study proves that availability, accessibility and responsiveness of DW are some of the critical features of the DW to attract end-users. Increasing the number of decision- makers, accessing the information by using different devices result in great impact on promotion of data ownership, which leads to informed-decision making process.

Data mining tools have to go further than just to provide an output. Data-mining tools should support the output that creates knowledge, and it is the knowledge that can be used to make decisions about several business problems [14]. The finding shows the pivot table and Data Visualizer are useful for more than 90%. The Evaluation result shows the usefulness of GIS is less than 80%, which is a big surprise. GIS is considered as modern way of presenting information to answer "where" there is concern. Awareness and training is required to enable the decision makers to learn and know knowledge produced by GIS and yet they are missing. DW with competency data mining and presentation tools, which work collaboratively with OLAP to fetch and present data into different dimension and format, easily interpreted by decision makers supports the informed decision-making. DW that present data in user-friendly and on- demanded aggregation basis, improve the data usage in organization [4]. Besides, it improves the quality of data from HIS subsystem "the more data are used, the more data quality will improve" [20] through comments and feedback that will raise accountability of sub-systems. As a result it even will help dispel the perception that data collected by routine HIS sub-system has low quality.

Successfully implemented, DW must have competency BI tools. Results revealed that Dashboard and Score card are at the heart of the modern DW and interests of decision- makers in health sector. This concurs statement from Coronel, Morris & Peter Rob "BI is about creating intelligence about a business, this intelligence is based on learning and understanding the facts about a business environment" [14]. The intelligence generated by BI is important in fostering informed decision-making. BI tools displaying a set of irrelevant glossy graphics full of attractive images without a vivid message, is unable to aid in decision making at health sectors. Therefore, BI tools should be tailored in a way that best reflects business environment and communicates due message to decision makers.

Dashboard needs to present information focusing at decision maker's management hierarchies; presenting data in appropriate presentation format (e.g. Pivot, graph, GIS) with real disaggregation are more helpful in fostering decision-making. Furthermore, DW in health sectors allows eligible end-users to create their own multiple pages of dashboard by subject to follow-up. The scorecard is getting more interesting and greater advocacy in health sector as time goes on. The

scorecard promotes informed decisions due to its in-built capability - in creating clear communication with end-users, depicting comprehensive- but easy to interpreting (what). This helps creates the overall picture on the performance of health sector, sets strategic priority of Health Programs clearly as the color codes are set, based on milestones and targets set by the sector.

A. Gaining the trust of users is real key to success in DW

It is important that if organization wants to sustain DW, overall trust can be built in terms of data security, data quality and data sharing from all possible sources. The end-users needs confidence or otherwise, DW, in place, won't tempt decision makers to utilize it. Sadly, the World of technology is full of people with malicious intentions and all online systems are vulnerable to security breach. Therefore, DW needs to be viably secured by mitigating the security risks as best as possible. Unlike personal information, which needs high level of confidentiality, security and privacy, DW stored aggregated data are less sensitive. The study shows that more participants are satisfied with security on data on transit (online communication encryption using TLS), data at rest (protection and monitoring of server by using firewall) and end-user security (user access control or user role).

(a) Data Sharing: The DW is not source of data; it just depends on data from other systems; therefore the systems need trust before sharing the data to DW. DW means more users will access the information that organizations need to ensure that the ethical standard is duly applied. Authorized users utilize data for authorized reasons and should not disclose any information to a third part without consent of organizations.

(b) Data quality: The quality of data is an essential component for success of any DW. Data quality in DW depends entirely on the quality of data from sources. If the data from the source is poor, the classic notion of garbage-in garbage-out applies. Apart from the sources, ETL is an area that can cause poor quality of data. According to researchers [21], this area is challenging, has few researches, and it has no standard model to represent ETL.

B. Challenges and Risks

The participants mentioned that there is minor external dependency risks, but the solution is open source. Most computer science graduates in Zanzibar have capacity to handle the system. There are potential technical and human resources that can manage the DW, which is based on open source knowledge. With proper training, capacity- building programmes the local team can support the developments aspects in case there is no support of DHIS2 software. However, participants proposed training for more people in order to mitigate risks of sustaining the project in term of maintenance and improvement in future. Availability of fund to support implementation and maintenance of project, timely availability and quality of data collected at the sources, limited technical and operational champions in the sector and lack of top managerial project ownership in organizations in encouraging the use of DW account for more than 50%.

C. Recommendations / Prospects for Future Research

The research project applied ETL model, which is not sophisticated (not fully automated). This has a risk on data quality as data can be tampered along the process. This area is important and it needs further research in future. The task may need close cooperation from each system administrator's managing operational databases. Further in-depth study on interoperability and security could be conducted on Health Level Seven International (HL7) and Statistical Data and Metadata Ex-change - Health Domain (SDMX-HD). These are international initiatives to simply demand for health system interoperability.

The DW is relative a new idea in Zanzibar health sector. The universities and organizations should take further steps to rethink their curriculum, conduct seminars, workshops, and meetings on DW in supporting modern informed management. There is strong needs to introduce the unique identifier code for organizational hierarchy i.e. health facilities, if there is any existing unique code for Health facilities or data elements, The sector needs to manage the due integration process by adding modification to the data sources system to simplify the interoperation factor.

VI. CONCLUSION

DW is useful in promoting the decision- making process in health sector but technology is not everything in development of informed decision-making culture in organizations. Non-technical interventions to work collaboratively with DW are required. These include incentives modality for employees or levels in organizations, performance- based financing (PBF), due accountability and regular sector performance assessments, performance recognition at different level of organizations. Effective strengthening of analytical capacity and interpretation skills are considered as collaborative intervention in promoting informed decision-making. It is important to note that development of DW needs data from different sources: therefore, the poor culture of sharing data between the units and departments, Health programmes and development partners are challenges that need to be discussed prior to adoption of implementation of DW. It is quite obvious that the heath sector is very rich in papers (well- documented policies and strategies, diverse domain of health Indicators with its composition clearly defined, categorized and elaborated). DW and BI tools are indeed proper mechanisms to transform from the office shelves to well- fashioned automated system.

REFERENCES

[1] Tekabe Belay, NkosinathiMbuya, &VikramRajan 09), Data Utilization and Evidence-Based Decision Making in the Health Sector (Survey of Three Indian States), Constella Futures group Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955.

[2] Swain Scheps (2007), Business Intelligence for Dummies, Wiley Puplishing, Inc.

[3] Tara Nutley (2012), Improving Data Use in Decision Making, An Intervention to Strengthen Health Systems, MEASURE Evaluation, Special reports.

[4] MEASURE Evaluation (2011), Tools for Data Demand and Use in the Health Sector.

[5] K.W. Chau, Ying Cao, M. Anson, Jianping Zhang (2002), Application of data warehouse and Decision Support System in construction management, Automation in Construction 12 (2002) 21, 3–224, Elsevier Science B.V.

[6] JørnBraa, Arthur Heywood, SundeepSahay (2012), Improving quality and use of data through data-use workshops: Zanzibar, United Republic of Tanzania, Bulletin of the World Health Organization,90:379-384.

[7] Seyed Ali MussaviRizi, Abdul Roudsari (2013), Development of a Public Health Reporting Data Warehouse: Lessons Learned, IMIA and IOS Press.doi:10.3233/978-1-61499-289-9-861.

[8] Matthew Grob, David Hartzband (2008), Health Centers and the Data Warehouse, Information Bulletin #14, HEALTH CENTER CONTROLLED NETWORKS SERIES (HCCN), National Association of Community Health Centres, Inc.

[9] Irtishad Ahmad, Salman Azhar, PranasLukauskis (2004), Development of a decision support system using data warehousing to assist builders/developers in site selection, utomation in Construction 13 (2004) 525–542, Elsevier B.V.

[10] S.Saagari, P.DeviAnusha, Ch.LakshmiPriyanka and V.S.S.N.Sailaja (2013), Data Warehousing, Data Mining, OLAP and OLTP Technologies Are Essential Elements to Support Decision-Making Process in Industries, International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-2, Issue-6.

[11] James Craver, Robin Pendley, Kate Brett, Li-Hui Chen (2015), Health Indicators Warehouse: Web Access to Population Health Indicators, Available http://www.cdc.gov/nchs/ppt/nchs2015/Pendley_Monday_BrooksideAB_C1.pdf. National, Conference on Health Statistic.

[12] WHO(2010) , Everybody's business — Strengthening health systems to improve health outcomes. WHO's framework for action. Geneva, World Health Organization, 2007

[13] Tanty Oktavia (2014), Implementing Data warehouse as a foundation for decision support system (perspective: technical and nontechnical factors), Journal of Theoretical & Applied Information Technology. Vol. 60 Issue 3, p476-482.

[14] Carlos Coronel, Steven Morris & Peter Rob (2011), Database Systems: Design, Implementation, and Management, Cengage Learning, 9th Edition.

[15] Edward T. Chen (2012), Implementation Issues of Enterprise Data Warehousing and Business Intelligence in the Healthcare Industry, Communications of the IIMA, 39 2012 Volume 12 Issue 2

[16] Azwa Abdul Aziz, Julaily Aida Jusoh, Hasni Hassan, Wan MohdRizhan, Wan Idris, Addy Putra MdZulkifli, ShahrulAnuwar Mohamed Yusof (2014), A Framework for Educational DWH(EDW) Architecture using Business Intelligence (Bi) Technologies, Journal of Theoretical and Applied Information Technology 10th, November 2014. Vol. 69 No.1

[17] Balanced Scorecard Institute -BSI (2016), [Online] Available at: http://balancedscorecard.org/Resources/About-the-Balanced-Scorecard.

[18] Salvatore T. March, Alan R. Hevner (2005), Integrated decision support systems: A data warehousing perspective, Decision Support Systems 43 (2007) 1031– 1043, Elsevier B.V

[19] Zanziber, Health Information Survey, 2011.(http://www.who.int/healthsystems/strategy/everybodys_ business.pdf).

[20] Shaker H. Ali El-Sappagh, Abdeltawab M. Ahmed Hendawi, Ali Hamed El Bastawissy (2011) A proposed model for DWHETL processes, Journal of King Saud University – Computer and Information Sciences 23,91–104